DOCUMENT RESUME

ED 146 210 TM 006 599

AUTHOR Gabriel, Roy M.
TITLE Effects of Single Subject Incomplete Block Designs on Nonmetric Multidimensional Scaling Solutions.
PUB DATE Apr 75
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Data Analysis; *Data Collection; *Mathematical Models; Matrices; *Multidimensional Scaling; Statistical Analysis
IDENTIFIERS *Incomplete Designs

ABSTRACT

Multidimensional scaling (MDS) a highly reliable measurement technique, often requires an overwhelming task of the subject in the data collection procedure. This investigation was designed to determine the loss of precision in solution associated with five degrees of systematic reduction in the data collection task. Data were simulated via Monte Carlo methods previously employed in comparable MDS investigations and represented a partially balanced incomplete block design for each subject. Although a linear decrease in solution precision with design efficiency was evidenced, these losses were nearly always within tolerable limits and had little effect on the interpretation of results. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED146210

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Roy M. Gabriel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Effects of Single Subject Incomplete Block Designs on Nonmetric Multidimensional Scaling Solutions

Paper presented at the 1975 Annual Meeting
of the American Educational Research Association

Roy M. Gabriel
University of Manitoba
April, 1975

TM006 599

With the many advances realized in the theory and practice of multidimensional scaling (MDS) as recently as the past fifteen years, it is difficult to present a general description which encompasses the wide variety found in its analysis algorithms and its applications. Still, as an introduction to the content of this paper, the following should suffice. Given a set of stimulus objects which differ from each other on several attributes or qualities, multidimensional scaling attempts to accomplish two general tasks: to solve for the number of dimensions along which the stimuli vary, and to solve for the scale values of each stimulus object on each of these dimensions. In more mathematical terms, the spatial configuration of a set of k stimulus objects in p-dimensional space is obtained. The reference axes which represent these dimensions may be rotated and subsequently interpreted in a manner akin to that of factor analysis.

The task set before the subject in an MDS investigation is invariably one of interstimulus comparison. Its purpose is to determine the extent of perceived distance between the stimulus objects. This distance is then utilized in the solution for a more metric expression of the distance between stimuli, and a suitable dimensionality is found to account for these distances to establish the aforementioned spatial configuration of the set. Scale values for each stimulus on each dimension are merely the projections of the point in space representing the stimulus object onto the reference axes.

As mentioned, variations of nearly every feature of MDS contained in this general overview are in abundance. For example, the decision made by the subject regarding the stimuli before him may be in terms of their similarity (or dissimilarity), substitutability, co-occurrence, or preference. Marketing research often employs the preference and substitutability criteria in judging various manufactured products (breakfast foods, automobiles, etc.) to determine consumer perceptions.

Apart from these differences in decision criteria there exist a variety of data collection methods, distance models, and analysis algorithms. This paper does not attempt a comprehensive review of current methods. Such a taxonomy has only recently been compiled in a two-volume work by Shepard, Romney, and Nerlove (1972).

Specific Objectives of the Study

Multidimensional scaling seems to hold great promise as a measurement and analytical device for determining the perceptual space of a sample of subjects regarding a given set of stimulus objects. As the number of stimuli grows large, however, the task set before the subject, regardless of the particular method of data collection (pairwise comparisions, complete triads, etc.) may rapidly become unwieldly. The nature of the stimulus objects themselves heavily influences this condition. For example, rendering the 190 judgments of similarity (dissimilarity, preference, etc.) among pairs of twenty color chips or twenty polysyllabic adjectives is probably far less taxing than the same task applied to twenty student profiles or twenty short stories. Still, the fatigue factor in operation during the task may inject serious error into the procedure and distort the eventual solution.

The intent of this investigation was to simulate a method of systematically reducing the task for each subject by requiring of him only a fixed proportion of all possible pairwise judgments contained in the design. Complete similarity data matrices were generated via the Monte Carlo procedure discussed in the following section for various combinations of sample size (n), number of stimulus objects (k), and dimensionality of true solution (r). These were then independently subjected to five levels of data elimination. The remaining similarity judgments were aggregated across the sample and the aggregate similarity matrix submitted to MDSCAL (Kruskal & Carmone, 1969) for analysis.

The worth of the various fractional solutions was judged on three criteria:

1) Stress--goodness of fit of the MDSCAL solution to the similarity ordering.

2) Metric Determinacy--squared correlation between the "true" interstimulus distances and those of the MDSCAL solution.

3) Stress by Dimensionality plots for the five fractional solutions in a given $\underline{n}$ x $\underline{k}$ x $\underline{r}$ combination.

The product of the simulation procedure employed seems analagous to item and matrix sampling techniques utilized in criterion referenced testing (Lord, 1962). In this regard, a word must be said about a possible "context effect" as it applies to multidimensional scaling. Do stimulus pairs occupy the same position in perceptual space when viewed in combination with only a portion of other pairs than they would if viewed in the complete data context? Although incapable of scrutiny in this investigation, since the data were simulated in the complete data context, there is some evidence for the tenability of this assumption. Research into the existence of a context effect in item sampling generally discounts the existence of a serious context effect under typical sampling circumstances. Though not the focus of their MDS investigation, Green, Maheshwari, and Rao (1969) split the 27 stimulus objects they used into two sets of sixteen which contained five stimuli in common. Analysis results were virtually identical in terms of dimensionality, its interpretation, and the scale values of the common stimuli. This is not meant to provide conclusive evidence, but rather an indication that for the present, overlooking the context effect in this investigation is reasonable.

Previous Research on Incomplete Data Matrices in MDS

Young and Cliff (1972) devised a system for data collection which drastically reduces the task for the subject while still involving a large number of stimulus objects in the experiment.

Their Interactive Scaling for Individual Subjects (ISIS) cuts down on the number of judgments necessary through intermittent analyses of small portions of the collected data performed while the subject is carrying out the task. His responses are stored and immediately analyzed via an on-line computer setup. The results of the analysis are used to begin to frame his perceptual space and to determine which stimulus pairs he views in the next sample. This process continues until non-significant improvement in the multidimensional scaling solution is attained. The point of diminishing returns was typically reached with fewer than one third of the possible judgments contained in the set of stimuli, although the authors employed a large number of stimulus objects (k = 43) to demonstrate the method's capability of scaling a large set of stimuli without requiring an unreasonable amount of judgments from the subject.

The sequential nature of the ISIS algorithm sets it apart from the others currently in use for multidimensional scaling analysis. Neither its appropriateness nor the stability of its results are questioned here. Its feasibility as an alternative method of data collection and analysis for multidimensional scaling investigations may be limited, however. Both the hardware and software support needed for its application are considerable and probably not conveniently available for the practicing researcher. Furthermore, the efficacy of the algorithm is partially a function of the appropriateness of the pairs of stimuli contained in the first sample shown to the subjects. If the experiment is exploratory in nature, the insight necessary for this step may not be readily at hand and the subsequent reduction of the task not as dramatic.

Domoney, Spence, and Ogilvie (1973) simulated the use of incomplete designs in nonmetric multidimensional scaling and found the metric recovery to be "satisfactory" regardless of error level when the "degrees of freedom Ratio" exceeded 3.5.[1] Their Monte Carlo simulations did not generate individual subject

---

[1]For k stimulus objects and an r dimensional solution this degrees of freedom Ratio is given by: $\frac{k(k-1)/2}{k(r-1) - r(r-1)/2}$

data, but only the aggregate similarity matrices. Thus, unlike the investigation to follow in this paper, the incomplete data in the Domoney et al(1973) study was in the aggregate data matrix. Their elimination procedures included both random deletion of a fixed proportion of the data and the application of cyclic incomplete block designs of various levels of efficiency. They found the latter to provide slightly better metric recovery, though not significantly in most cases. Their simulations were of 32, 40, and 48 stimulus objects in three dimensions with various levels of error perturbation. They advised caution in generalizing their results (particularly the 3.5 value of the degrees of freedom Ratio beyond which high levels of error did not significantly retard metric recovery) to smaller sets of stimulus objects, and called for more research under those conditions.

METHOD

For the remainder of this paper, the notational conventions proposed by Millman and Glass (1967) will be followed. Capital letters will be used with reference to a given parameter, while lower case letters will be used to refer to a specific level or levels of that parameter. The notation used and the levels represented of the parameters under investigation in this study are given below:

Parameter	Symbol	Levels
Number of stimulus objects	K	15, 20
True dimensionality	R	2, 3
Sample size	N	20, 40
Proportion of judgments remaining for each subject (efficiency)	E	1.00, .80, .60, .40, .20

A "true" perceptual configuration of the k stimuli in r dimensions was established for each of the four combinations of k and r. In any one of these four combinations the kr coordinates were generated in a manner identical to that employed by Sherman (1972). In each of r dimensions k coordinate values were randomly generated from a uniform distribution on the interval - 1.00 to + 1.00. This procedure results in a random scatter of points in a given dimensionality. This in no way implies a lack of structure or interrelationship in the data, for the specification of true dimensionality is the structural influence as far as the MDS algorithm is concerned. In other words, the algorithm seeks only the dimensionality which best reconstructs the interstimulus distances (represented by the presumably monotonically related similarity judgments) in solving for the scale values of the stimuli on these dimensions. In the absence of error variation, a solution in the appropriate true dimensionality would experience zero Stress ("perfect" fit) in a very few iterations. The four similarity matrices corresponding to these true configurations (i.e., no error) were submitted to MDSCAL to verify this principle and all reached zero Stress within three iterations.

Unless an individual differences model is prescribed, the assumption governing each subject's perceptual space is that it differs from the true perceptual space due to random error only. Thus, prior to computing the interstimulus distances and monotonically related similarity data for each subject in the sample, the true positions of the k points in r dimensional space were perturbed by random error components unique to each subject. This error perturbation represented the introduction of between subject variation in their perceptual spaces of the k stimulus objects in r dimensions. The magnitude of this error injection was determined via the E ratio suggested by Young (1970), but improved by Sherman (1972). The variance of the normally distributed error deviates was proportional to the variance of the coordinate values in each dimension of the true configuration.

$$E = \frac{\sigma_{er}}{\sigma_{cr}}, \qquad (1)$$

where

σ_{er} = standard deviation of error components on dimension r, and

σ_{cr} = standard deviation of coordinate values on dimension r.

To provide a workable degree of intersubject variability the value of .50 was chosen for E. The variance of the error deviates was thus one fourth as large as the variance of the true coordinate values.

Following this intersubject variance introduction, a second injection of error in the data took place. This was within subject error, representing the internal inconsistency in the data collection task evidenced by each subject. As described by Young (1970) and replicated by Sherman (1972) the appropriate point at which to inject this error is just prior to each interstimulus distance computation. For example, to compute the distance between two points in two dimensions, all four coordinates were perturbed by adding a random normal error deviate according to the formula:

$$x'_{ir} = x_{ir} + e_{ir(j)}, \qquad (2)$$

where

x'_{ir} = error-perturbed coordinate value of stimulus i on coordinate r, and

x_{ir} = true coordinate value of stimulus i on coordinate r, and

$e_{ir(j)}$ = error component added to stimulus i on dimension r in the j^{th} distance computation $[e_{ir(j)} \sim N(0, \sigma^2_{er})]$.

This procedure was repeated for each of the $k(k-1)/2$ distance computations. In other words the position of a stimulus object in the r dimensional space was perturbed independently for each of its $k-1$ distance computations. Aside from insuring a non-trivial solution (c.f., Young, 1970), in the author's opinion this better represents the judgmental process of the subject, as it allows for a shifting of a given stimulus object's position in his perceptual space each time it is paired with another stimulus. The magnitude of this intrasubject error was again determined using the E ratio given in (1). Its value was fixed at .35 for each dimension.

Although this level of intrasubject variability was initially chosen as one which provided a visible difference in goodness of solution in both Young's (1970) and Sherman's (1972) Monte Carlo investigations, it has interpretable meaning in terms of each subject's internal consistency in the interstimulus judgmental tasks. If the "true" coordinate values of the stimuli are considered true scores and the error deviates considered error scores, the internal consistency or reliability of any individual's judgments is

$$\rho = 1 - \left(\frac{\sigma_e}{\sigma_t}\right)^2 \tag{3}$$

The choice of E = .35 therefore corresponds to an intrasubject reliability of .89 as shown below.

$$\rho = 1 - (.35)^2 \tag{4}$$

$$\rho = .89 \tag{5}$$

Although this may appear an unrealistically high standard, both Shepard and Kruskal have pointed out the probable high reliability associated with such highly redundant tasks as those involved in multidimensional scaling. As part of the demonstration of the efficacy of their ISIS procedure,
Young and Cliff (1972) showed thirty of the stimulus pairs to

their subjects twice. In this way they obtained a measure of internal consistency for each subject. The mean correlation of the similarity ratings of the thirty pairs across subjects was .909.

With the error pertubed distances computed, only one step remained for generation of each subject's complete data matrix--the monotonic transformation of those distances to the similarity (or dissimilarity) judgments representing the product of typical MDS data collection procedures. As has been abundantly noted by most of the previously cited authorities in the area of multidimensional scaling, the particular transformation employed is unimportant and has no effect on the goodness of fit of the solution. In this investigation, the following transformation, a normal density function utilized by Shepard (1962b) was employed:

$$s_{ij} = e^{-1.4\, d_{ij}^2} \qquad i \neq j \qquad (6)$$

where

s_{ij} = measure of similarity between stimuli i and stimulus j, and

d_{ij} = computed distance between stimulus i and stimulus j.

At this point in the procedure, the complete data matrices had been generated and only the elimination procedure to be discussed shortly was needed for the data to be submitted for analysis. A summary of the steps in the complete data generation is given below.

1) K, the number of stimulus objects in the configuration was specified. K took on values of fifteen and twenty.
2) R, the dimensionality of the true perceptual space was specified. R took on values of two and three.
3) The kr coordinate values were randomly generated from a uniform distribution on the interval - 1.00 to + 1.00.

4) Intersubject variability was introduced using .50 as the value for E (1). Random error deviates $\sim N(0, \sigma^2_{er})$ were added to each coordinate value in the true perceptual space to produce a randomly discrepant perceptual space for each subject.

5) Intrasubject variability was introduced using .35 as the value for E (1). Random error deviates $\sim N(0, \sigma^2_{er})$ were added to each coordinate value in the subject's perceptual space and the $\underline{k}(\underline{k} - 1)/2$ interstimulus distances were computed via the Euclidean distance function.

6) The error perturbed distances ($\underline{d}_{ij}$) were transformed to monotonically related similarities ($\underline{s}_{ij}$) via the transformation in (6).

These six steps were repeated sixty times (N = 20, N = 40) for each of four stimulus-dimensionality combinations.

Data Elimination

The process of eliminating data from these complete similarity matrices proceeded along the lines of a partially balanced incomplete block (PBIB) design viewing the $\underline{n}$ subjects as blocks and the $\underline{k}(\underline{k} - 1)/2$ stimulus pair combinations as treatment levels. Aside from the fact that each subject experiences the same proportion of the $\underline{k}(\underline{k} - 1)/2$ treatment levels, the distinctive feature of a PBIB design is that each treatment level (stimulus pair) is experienced an equal number of times by all subjects. The number of treatments viewed by each subject, (stimulus pairs), expressed as a proportion of the total number of treatment levels is termed efficiency, which expresses the extent of data elimination in the design.

The complete data matrix of each subject was subjected to random elimination of the appropriate proportion of data for five levels of efficiency (1.00, .80, .60, .40, .20), subject to the aforementioned PBIB design constraints. For example, consider a set of fifteen stimuli, the 105 pairs of which were

judged by twenty subjects (k = 15, n = 20). For E = .8, the following represent the two demands of a PBIB design:

1) For each subject, 21 of the 105 data values are discarded.
2) Each of the 105 data values are discarded 4 times across the entire sample.

Steps one and two were performed independently for each of the five levels of efficiency (although no elimination was indicated for efficiency = 1.00), and the procedure was replicated three times. That is, for a given level of efficiency, the appropriate proportion of data was discarded three times, the pairs comprising that discarded proportion of data being randomly and independently chosen each time, allowing three replications of each level of efficiency (excluding efficiency = 1.00).

Briefly, then, the data elimination procedure was composed of the following steps:

1) K, the number of stimuli was specified. It took on values of fifteen and twenty.
2) R, the dimensionality of the true space was specified. It took on values of two and three.
3) N, the sample size was specified. It took on values of twenty and forty.
4) E, the proportion of remaining data was specified. It took on values of 1.00, .80, .60, .40, .20.
5) For each of n subjects, the appropriate proportion of data (1.00 - e) was discarded, subject to PBIB design constraints.
6) Step five was repeated three times.

At this point the incomplete data matrices of the combination of K, R, N, Efficiency, and replications were aggregated across the sample and submitted to MDSCAL for analysis. Solutions were sought in four dimensions: r - 1, r, r + 1, and r + 2 to provide a reasonable range for visual inspection of the Stress by Dimensionality graphs. To address the local minimum problem,

the similarities were scaled twice in each dimensionality starting from two different initial configurations, following a procedure outlined by Kruskal (1964b).

RESULTS

The experimental design employed is given in Figure 1 as a four factor repeated measures design with three replications crossing Efficiency, the repeated measures dimension. Tables I and II contain the results of the repeated measures MANOVA (Finn, 1969; Poor, 1973) performed separately on the Stress and Metric Determinacy criteria. Although analyzing multiple criterion measures separately is rarely, if ever, recommended (Gabriel & Hopkins, 1974) the extremely low p-values in evidence in the results (c.f., Tables I and II) together with the small number of criteria (two) serve to reaffirm the confidence that stable differences do exist among the levels of the parameters included in this study.

As in both Young (1970) and Sherman's (1972) studies virtually all factors exhibit significant differences on both Stress and Metric Determinacy. Their findings regarding the effects of K and R were generally replicated here. As the number of stimulus objects increases in a given dimensionality, the Stress coefficient rises, indicating <u>poorer</u> fit (in the presence of error in the data). The significant K main effect ($p < .0001$) is in the direction of higher Stress for K = 20 stimuli. Metric Determinacy also evidences the significant K main effect, and in the hypothesized direction--as the number of stimulus objects increases, metric recovery <u>improves</u>. This finding corroborates

Stimulus Objects (K)	Dimensionality (R)	Sample Size (N)		Efficiency (E) 1.00	.80	.60	.40	.20
15	2	20	r_1 r_2 r_3					
		40	r_4 r_5 r_6					
	3	20	r_7 r_8 r_9					
		40	r_{10} r_{11} r_{12}					
20	2	20	r_{13} r_{14} r_{15}					
		40	r_{16} r_{17} r_{18}					
	3	20	r_{19} r_{20} r_{21}					
		40	r_{22} r_{23} r_{24}					

TABLE I

MULTIVARIATE ANOVA OF REPEATED MEASURES FOR STRESS

Source of Variation	SS	df	F	p
K	.0416	1	164.70	.0001
R	.0032	1	12.62	.0027
N	.0857	1	339.09	.0001
E	.6594	4	1455.90	.0001
KR	.0134	1	53.02	.0001
KN	.0049	1	19.25	.0005
KE	.0059	4	3.57	.0356
RN	.0034	1	13.29	.0022
RE	.0101	4	3.72	.0314
NE	.0026	4	15.65	.0001
KRN	.0005	1	2.02	.1749
KRE	.0059	4	3.30	.0450
KNE	.0006	4	.32	.8616
RNE	.0013	4	2.16	.1310
KRNE	.0011	4	1.18	.3654
S(KRN)	.0040	16		
SE(KRN)	.0402	64		
TOTAL	.8838	119		

TABLE II

MULTIVARIATE ANOVA OF REPEATED MEASURES
FOR METRIC DETERMINACY

Source of Variation	SS	df	F	p
K	.3495	1	13.26	.0023
R	.6383	1	24.21	.0002
N	2.3409	1	88.80	.0001
E	2.6295	4	90.49	.0001
KR	.1183	1	4.49	.0502
KN	.1484	1	5.63	.0306
KE	.2523	4	2.60	.0851
RN	.1294	1	4.91	.0417
RE	.0276	4	.11	.9765
NE	.0696	4	2.77	.0729
KRN	.0392	1	1.49	.2406
KRE	.1048	4	1.36	.3011
KNE	.1188	4	1.83	.1827
RNE	.0470	4	.31	.8662
KRNE	.1174	4	1.60	.2321
S(KRN)	.0422	16		
SE(KRN)	4.3445	64		
TOTAL	11.5177	119		

TABLE III

CELL MEANS OF STRESS (S) AND METRIC DETERMINACY (M)

K	R	N	E 1.00	.80	.60	.40	.20	
15	2	20	.1348 .9545	.1668 .9032	.2171 .8546	.3232 .8931	.3818 .8094	$\bar{S}$ $\bar{M}$
		40	.0969 .9649	.1209 .9594	.1364 .9567	.2569 .8766	.3355 .8150	$\bar{S}$ $\bar{M}$
	3	20	.1112 .9388	.1559 .9050	.1827 .8850	.2264 .8739	.3167 .8408	$\bar{S}$ $\bar{M}$
		40	.0922 .9515	.1202 .9329	.1722 .9189	.2018 .9173	.2764 .8752	$\bar{S}$ $\bar{M}$
20	2	20	.1759 .9416	.2153 .9190	.2540 .9130	.3040 .8825	.3982 .8987	$\bar{S}$ $\bar{M}$
		40	.1229 .9758	.480 .9660	.1787 .9571	.2113 .9598	.3233 .9568	$\bar{S}$ $\bar{M}$
	3	20	.1882 .9360	.2112 .8914	.2689 .8891	.3087 .8726	.3923 .8605	$\bar{S}$ $\bar{M}$
		40	.1449 .9589	.1706 .9454	.2030 .9217	.2371 .9161	.3150 .9120	$\bar{S}$ $\bar{M}$

that of Young (1970) in recognizing the rise of the Stress coefficient as the number of stimulus objects increases under the same conditions of dimensionality, error, etc. This paradoxical consequence was rendered a measurement artifact of the Stress coefficient) by Young (1970) as true metric recovery (i.e., Metric Determinacy) was observed to improve with an increase in stimulus objects.[2] Finally, an increase in true dimensionality evidenced better apparent fit (Stress), but poorer true fit (Metric Determinacy). The inverse relationship here is predictable in light of the value of the degrees of freedom Ratio (c.f., footnote 1) for the four K x R combinations included in this study. Generalization of the relationship between S and M under various conditions of dimensionality should be attempted only with reference to this degrees of freedom Ratio.

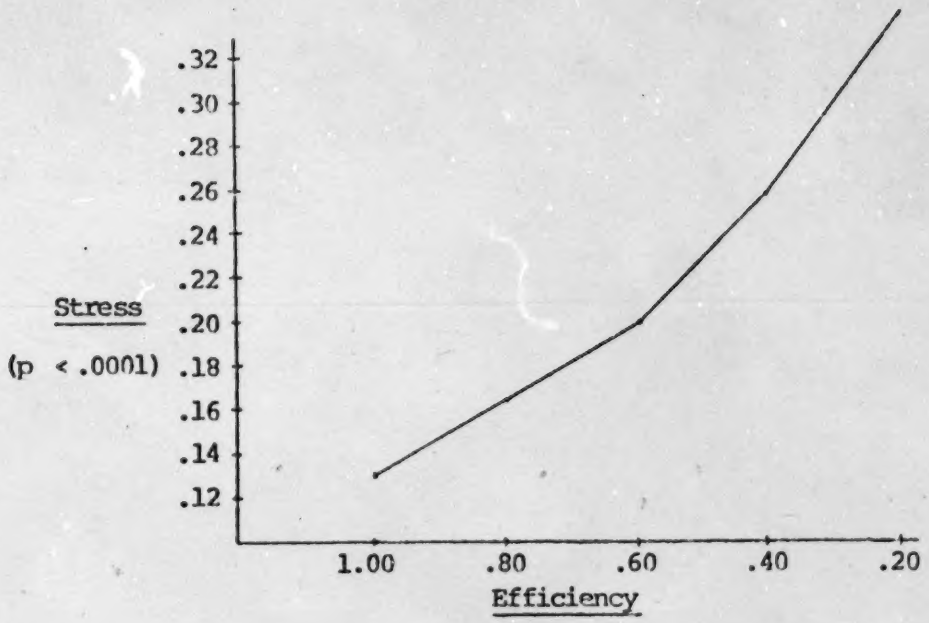
Sample size (N) was included as a factor in this investigation unlike other comparable studies and it evidenced consistently significant effects on the criteria. The N main effect was highly significant for both Stress and Metric Determinacy ($p < .0001$), with the larger sample size resulting in clearly superior solutions (lower Stress, higher Metric Determinacy). This is by no means a surprising outcome, but is an indication that sample size should be an important consideration in both simulated and applied MDS research.

The effects involving the Efficiency factor were the primary interest in this investigation. Figure 2 contains graphs of the E main effect for both Stress and Metric Determinacy. Highly significant differences among the levels of E were found in both criteria. Ninety-five percent confidence bands are included around values of M at each level of E in Figure 3 to lend some perspective to the statistical significance of differences across levels of Efficiency. The trend analysis for this E main effect revealed a highly significant linear trend

---

[2]The present author shares the widespread opinion that the Stress coefficient should be viewed as solely a descriptive statistic. Further inferential comments regarding the statistical analysis of the results of this investigation will be confined to the Metric Determinacy criterion.

Figure 2


Stress
(p < .0001)
.32
.30
.28
.26
.24
.22
.20
.18
.16
.14
.12
1.00
.80
.60
.40
.20
Efficiency

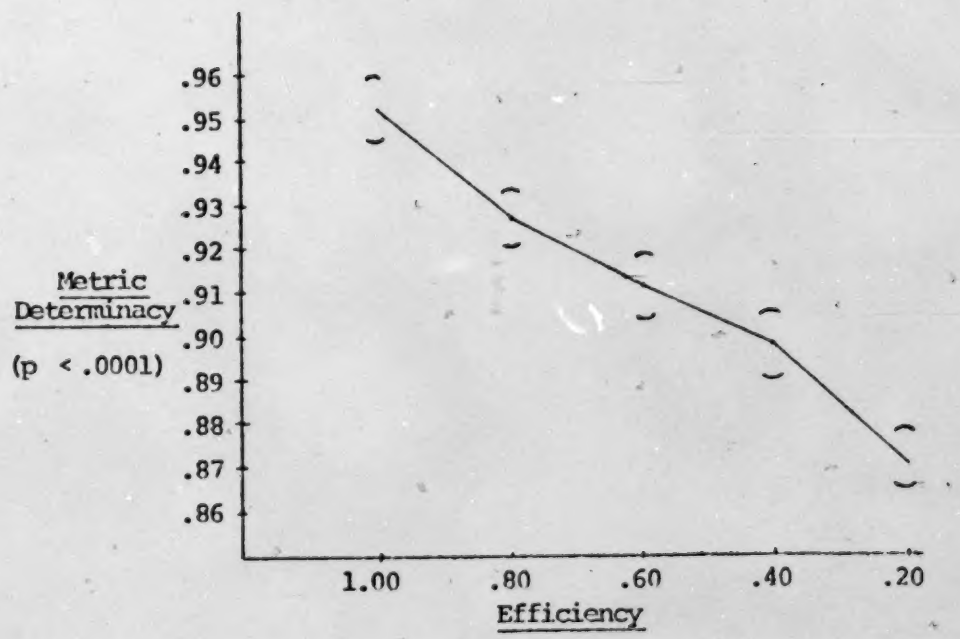

Metric
Determinacy
(p < .0001)
.96
.95
.94
.93
.92
.91
.90
.89
.88
.87
.86
1.00
.80
.60
.40
.20
Efficiency

TABLE IV

Source of Variation		SS	df	F	p
E		2.6295	4	90.49	.0001
	Linear	2.5204	1	53.76	.0001
	Quadratic	.0434	1	.49	.4956
	Cubic	.0596	1	.80	.3832
	Quartic	.0072	1	.12	.7375
K x E		.2523	4	2.60	.0851
	Linear	.1866	1	4.03	.0621
	Quadratic	.0477	1	.54	.4745
	Cubic	.0132	1	.18	.6792
	Quartic	.0025	1	.04	.8422
R x E		.0276	4	.11	.9765
	Linear	.0165	1	.35	.5609
	Quadratic	.0093	1	.10	.7503
	Cubic	.0001	1	.00	.9721
	Quartic	.0016	1	.03	.8744
N x E		.0696	4	2.77	.0729
	Linear	.0037	1	.08	.7831
	Quadratic	.0497	1	.56	.4655
	Cubic	.0050	1	.07	.7987
	Quartic	.0104	1	.17	.6860
K x R x E		.1048	4	1.36	.3011
	Linear	.0927	1	1.98	.1800
	Quadratic	.0101	1	.11	.7400
	Cubic	.0014	1	.02	.8900
	Quartic	.0004	1	.01	.9400
K x N x E		.1188	4	1.83	.1827
	Linear	.1011	1	.21	.6506
	Quadratic	.0795	1	.90	.3583
	Cubic	.0038	1	.05	.8234
	Quartic	.0256	1	.42	.5279
R x N x E		.0470	4	.31	.8662
	Linear	.0108	1	.23	.6374
	Quadratic	.0277	1	.31	.5847
	Cubic	.0002	1	.00	.9639
	Quartic	.0084	1	.14	.7173
K x R x N x E		.1174	4	1.60	.2321
	Linear	.0322	1	.69	.4197
	Quadratic	.0427	1	.48	.4982
	Cubic	.0312	1	.42	.5261
	Quartic	.0114	1	.18	.6730

($p < .0001$, c.f., Table IV) in Metric Determinacy, indicating poorer true fit with decreasing efficiency in the design.

Of the interactions involving the Efficiency factor, only the KE ($p < .0851$) and NE ($p < .0729$) sources of variation evidence even moderate significance. The KE interaction is characterized by an accelerated loss in metric recovery for the smaller set of stimulus objects ($K = 15$) beginning with the .60 level of Efficiency (cf. figure 3).

The NE interaction provides an interesting look at the relationship between sample size and design efficiency. Given the same total number of judgments in the sample, item sampling theory generally favors the large sample--small proportion of judgments per subject combination (Lord, 1962; Shoemaker, 1971; Barcikowski, 1972). That is, 50% of the judgments from 40 subjects provides better estimates of population parameters than does 100% of the judgments from 20 subjects. Admittedly, the two levels of sample size included in this investigation do not provide a very extensive look at this comparison, but there are two N x E combinations which amount to an equal number of judgments across the entire sample. These two are:

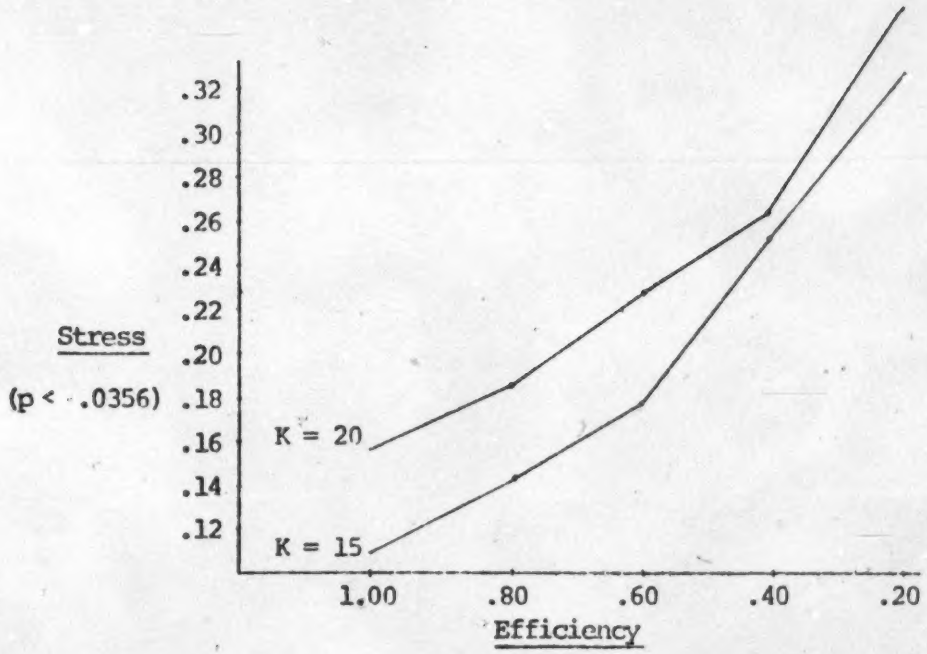
1) $\underline{n} = 20$, $\underline{e} = .80$ and $\underline{n} = 40$, $\underline{e} = .40$
2) $\underline{n} = 20$, $\underline{e} = .40$ and $\underline{n} = 40$, $\underline{e} = .20$

The horizontal dotted lines in figure 4 indicate that true metric recovery is better in the larger sample in both of the above situations. The discrepancy is only of the .01 magnitude, but its direction lends support to the item sampling theory expectation.

Other Criteria

Although the MANOVA clearly identifies the significant sources of variation among the parameters investigated here, these analytical results are meaningfully applied to Metric Determinacy alone--a criterion the user does not typically have at his disposal.

Figure 3


Stress
(p < .0356)
.32
.30
.28
.26
.24
.22
.20
.18
.16
.14
.12
K = 20
K = 15
1.00
.80
.60
.40
.20
Efficiency

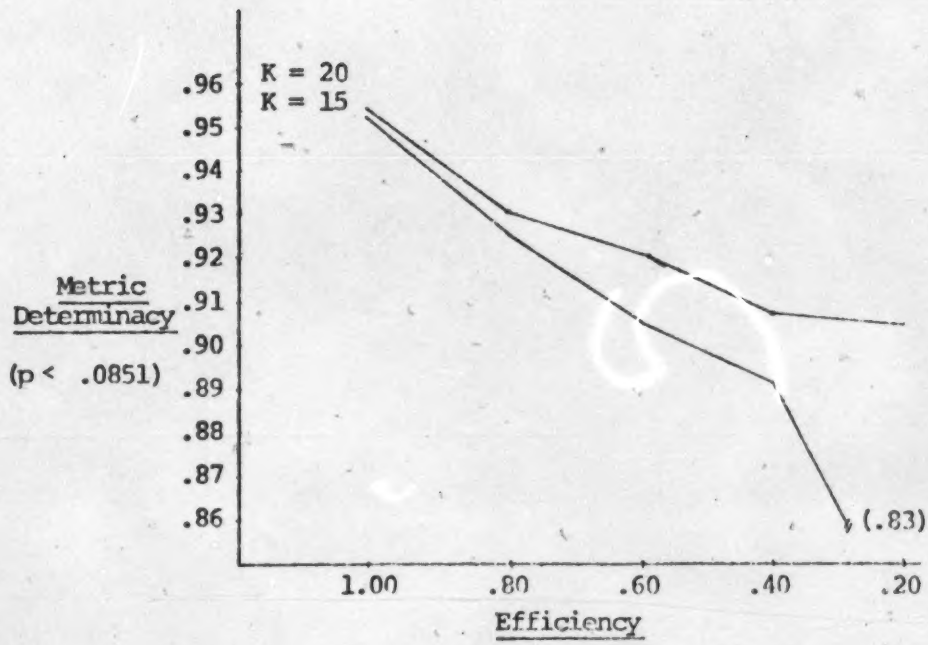
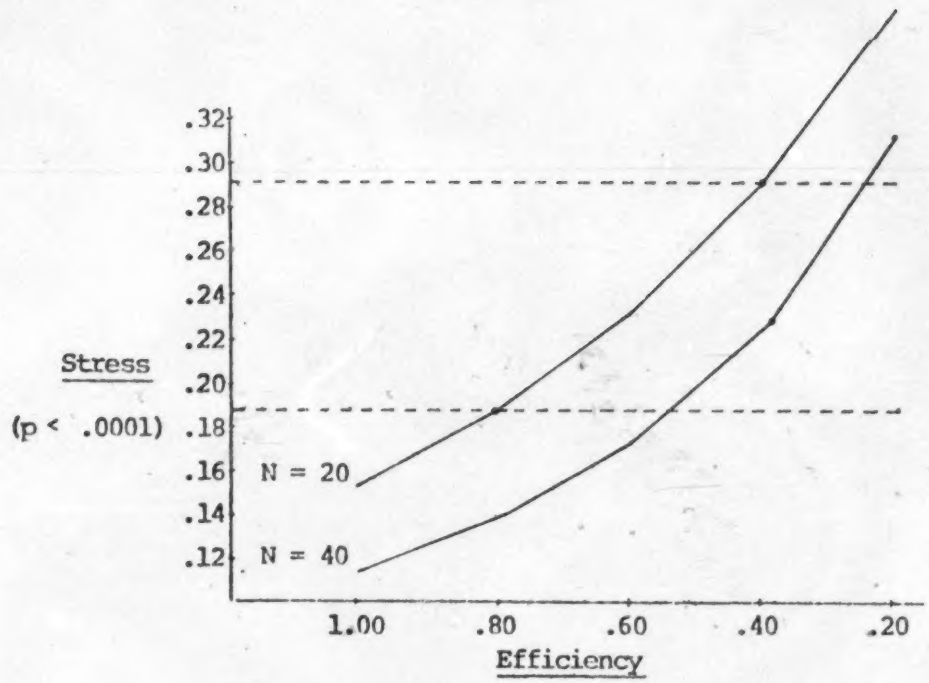
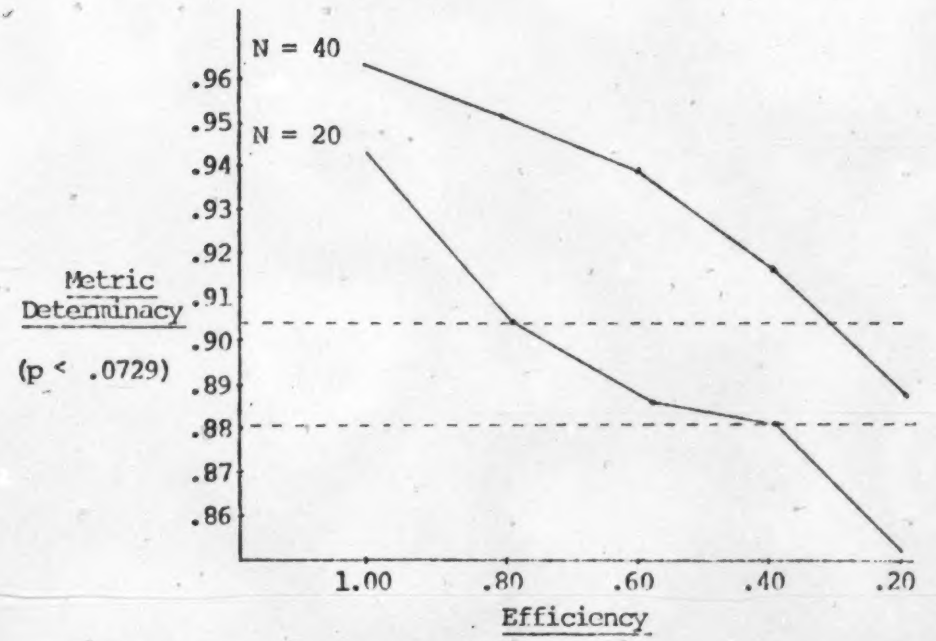

Metric
Determinacy
(p < .0851)
.96
.95
.94
.93
.92
.91
.90
.89
.88
.87
.86
K = 20
K = 15
(.83)
1.00
.80
.60
.40
.20
Efficiency

Figure 4


Stress
(p < .0001)
.32
.30
.28
.26
.24
.22
.20
.18
.16
.14
.12
N = 20
N = 40
1.00
.80
.60
.40
.20
Efficiency


N = 40
N = 20
Metric
Determinacy
(p < .0729)
.96
.95
.94
.93
.92
.91
.90
.89
.88
.87
.86
1.00
.80
.60
.40
.20
Efficiency

His task is to evaluate the pattern of Stress coefficients across the various dimensional solutions to find the conspicuous "elbow" in the series. Figures 5-12 display the Stress x Dimensionality graphs at each level of Efficiency for each of the eight n x k x r combinations. Although there are significant differences in Metric Determinacy across levels of Efficiency (c.f., Table II) along with obvious differences in Stress values in the expected direction, it is worthwhile to note that the choice of dimensionality based on these graphs is clear for all but one n x k x r combination. The "elbow" in the pattern of Stress values is indeed conspicuous at the appropriate dimensionality. Despite statistically significant differences in metric recovery of the various fractional solutions, the choice of dimensionality of solution is not ambiguous. With squared correlations on the order of .90 in nearly every case (c.f., Table III) these differences may not always take on meaningful proportions. The exception to this clarity in solution interpretation occurs when k = 15, r = 3, and n = 20 (c.f., figure 7). This coincides with the lowest degrees of freedom Ratio (c.f., footnote 1) and the smaller sample size. Although metric recovery remains quite good, it's clear that task reduction should not be attempted in this condition since the interpretive ambiguity may lead to the wrong choice of dimensionality and hence an inappropriate solution for his data. Note, however, that the larger sample size provides clear cut interpretation for the same k x r condition (c.f., figure 8). Again, sample size is seen as an important consideration.

DISCUSSION

In general, a high degree of metric recovery was found in the solutions even in low efficiency designs. Although M characteristically declined with efficiency, the loss in metric recovery

was not serious, in the author's opinion, particularly with the larger sample size. A loss of .07 in the squared correlation from E = 1.00 to E = .20 may well be within tolerable limits. The consistently high values for M seem to allow the researcher much flexibility in the choice of efficiency--sample size combination for his investigation. That is, if subjects are readily available, he may select a large sample and drastically reduce their task while not suffering a loss in metric recovery. If subjects are scarce, he may still obtain a precise solution by selecting a smaller sample and not reducing their task. There may be other priorities involved, such as the nature of the stimuli themselves. If they require a formidable amount of concentration, perhaps greater effort should be directed toward acquiring a larger sample to permit task reduction per subject. The salient point is that the findings of the present study indicate that the researcher may choose his combination of conditions with the expectation of minimal loss in the precision of the solution. Again, this is merely an expectation provided by these results. The author believes it important to corroborate them in an applied setting.

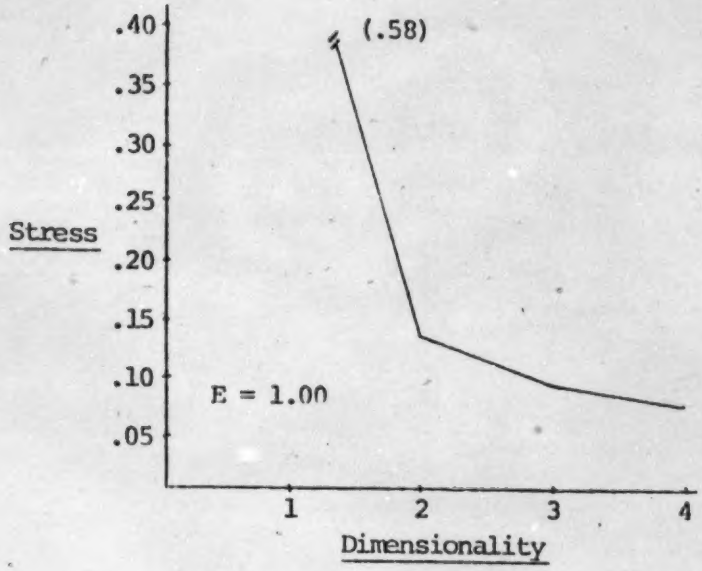
The previous comments were in terms of the Metric Determinacy criterion. Typically, this is not at the disposal of the user. True distances are obtained only from the true configuration and, unless the study is highly confirmatory in nature, this is not available. It was employed as a criterion in this and other simulated investigations because it measures the true goodness of fit of the solution. The apparent goodness of fit is measured by the Stress coefficient, a statistic influenced by many other factors, some of which were pointed out by Young (1970) and echoed by Sherman (1972). The correspondence between Stress and Metric Determinacy is not as strong as one would hope, particularly for smaller stimulus set sizes (Sherman, (1972)). This places the user, who has only Stress available to him, in a difficult position. When confronted with a Stress value of .25 (only "fair fit" in Kruskal's rules of thumb for Stress formula 2), he may be discouraged with the apparent inaccuracy

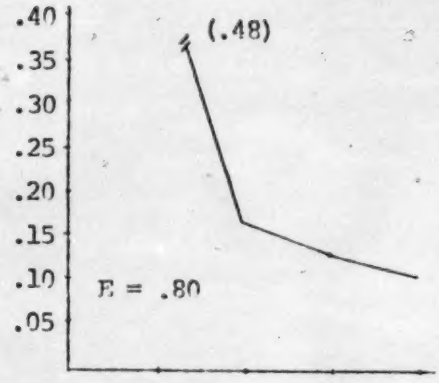
of solution. This solution may have a Metric Determinacy greater than .90 if the degrees of freedom Ratio is high, however. Surely this indicates a highly accurate recovery of distances. Due to this type of anomaly, the reliance on Stress should be strictly for determining the dimensionality of the solution, and not for evaluating the fit of the solution, in the author's opinion. Although fair to poor apparent fit was indicated by the Stress values associated with many of the fractional solutions in the present investigation, these values left the choice of dimensionality quite clear (c.f., Figures 8--15) in all but one k x r x n combination. This and other interpretive steps are the primary concerns of the user.

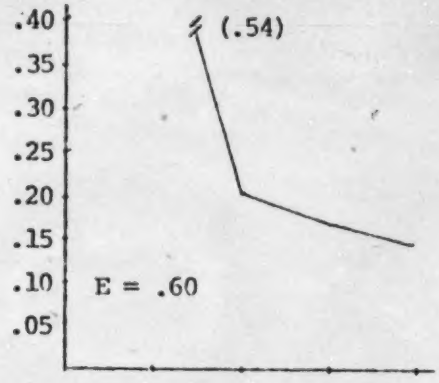
The author believes sample size to be a consideration worthy of further exploration. Again, to the practicing researcher, sample size is always an important concern. The results of the present study clearly show consistently more precise analyses for n = 40. Further, they provide an indication that large samples with small proportions of data collected provide as good and slightly better solutions than do small samples with large proportions of data. These initial findings are encouraging, especially in terms of the flexibility in planning they allow for the researcher.

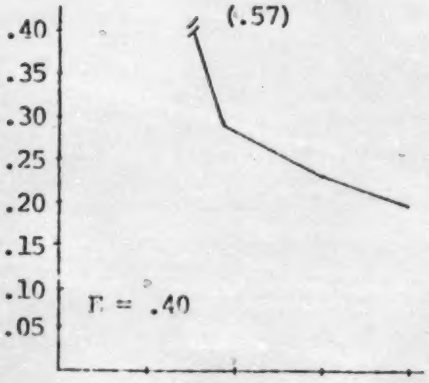
Figure 8

k=15 r=2 n=20









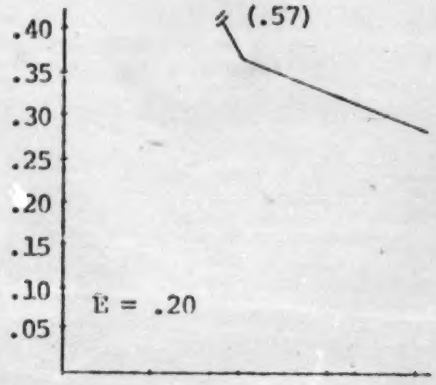
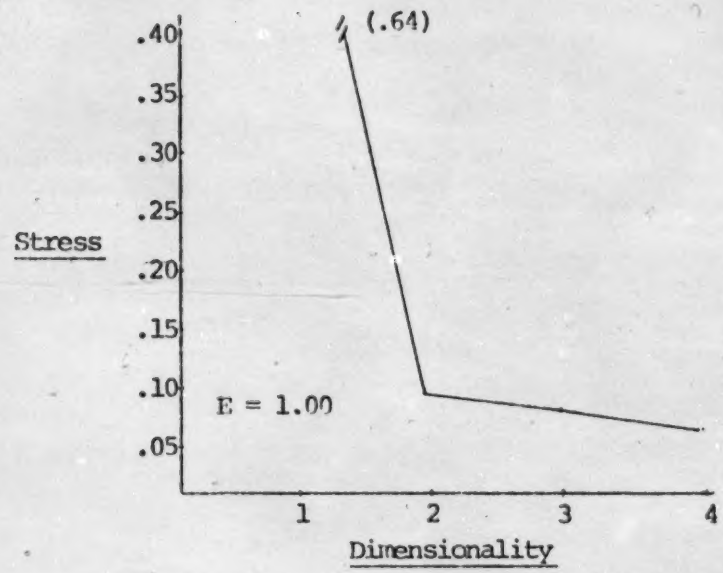
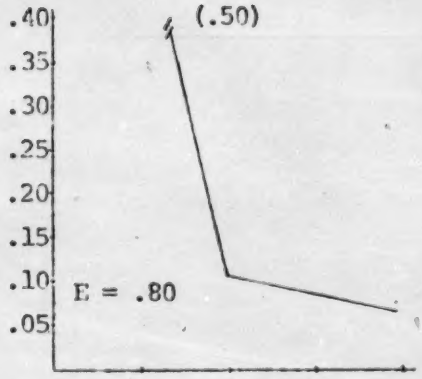
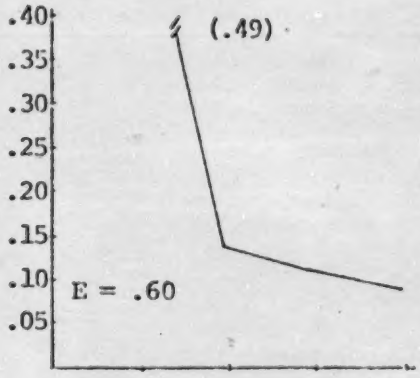


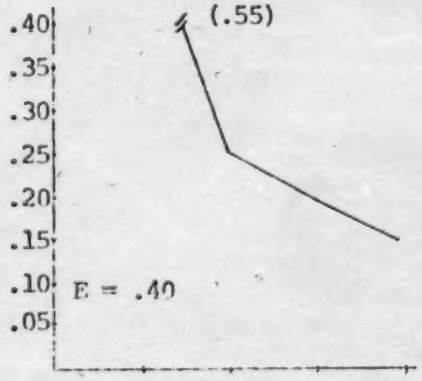
Figure 9

k=15 r=2 n=40









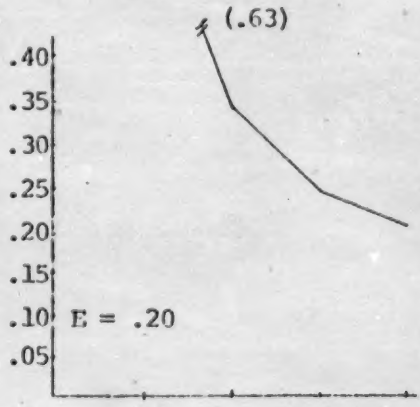
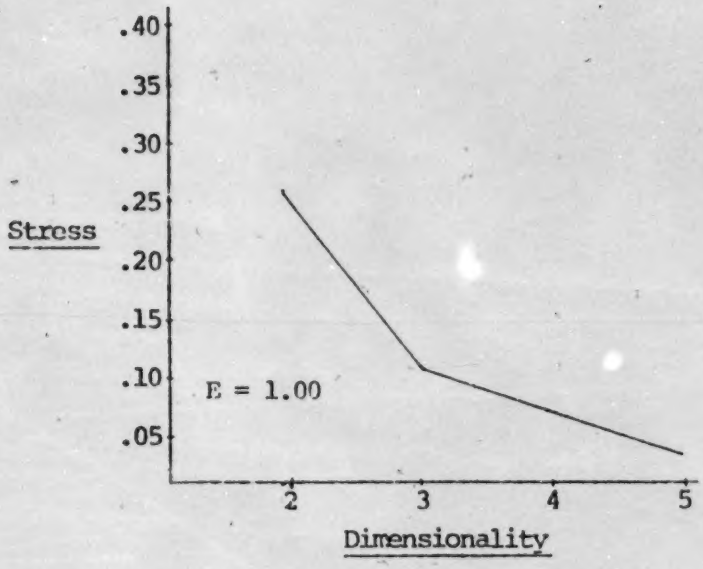
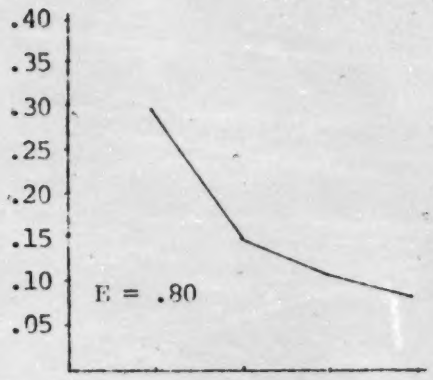
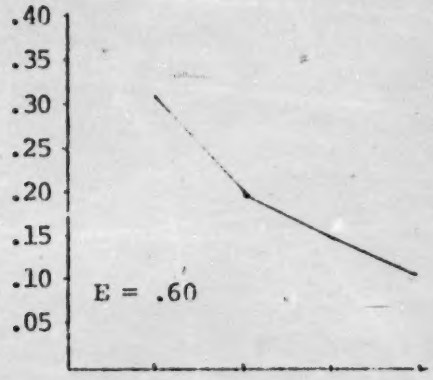


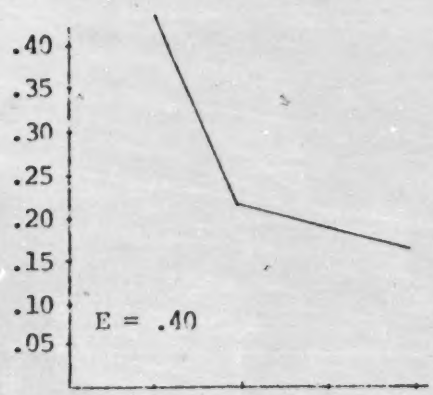
Figure 10

k=15 r=3 n=20









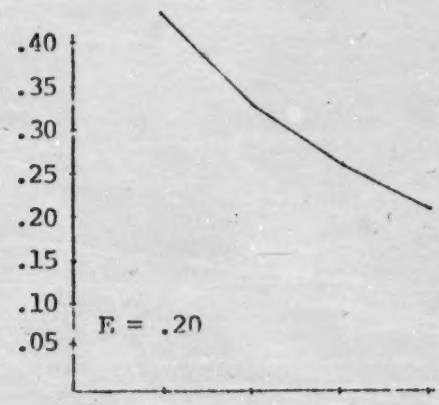
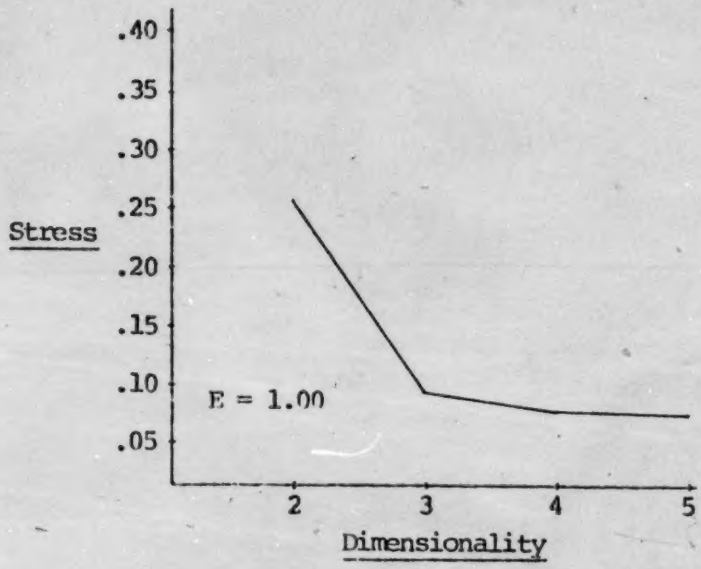
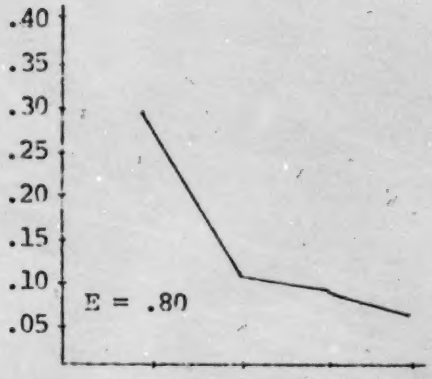
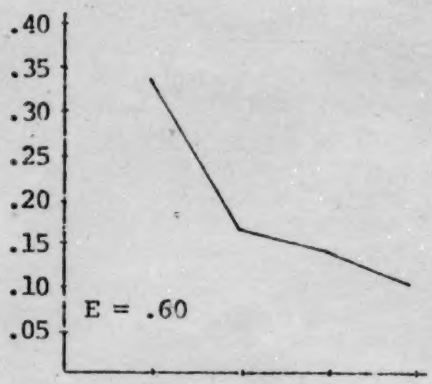


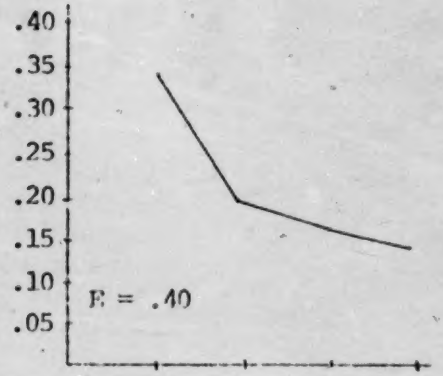
Figure 11

k=15 r=3 n=40









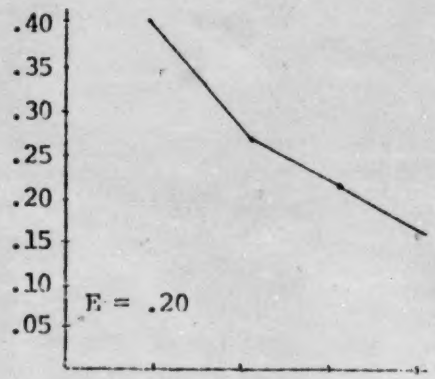
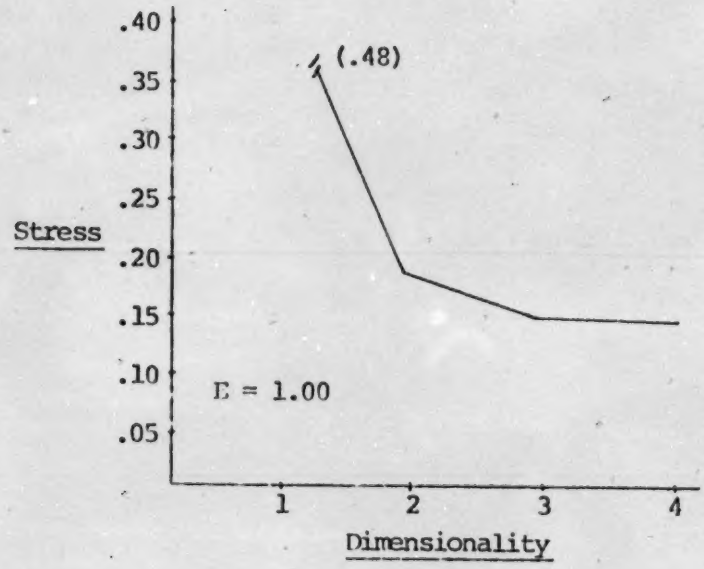
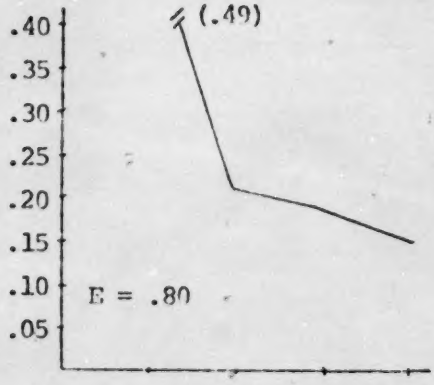
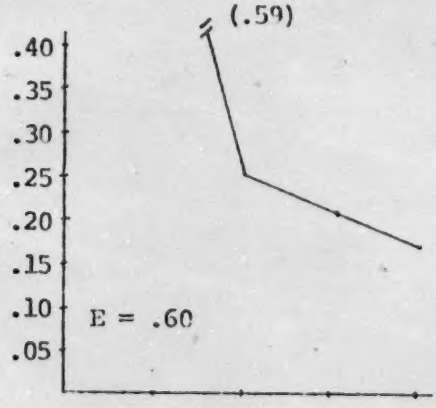


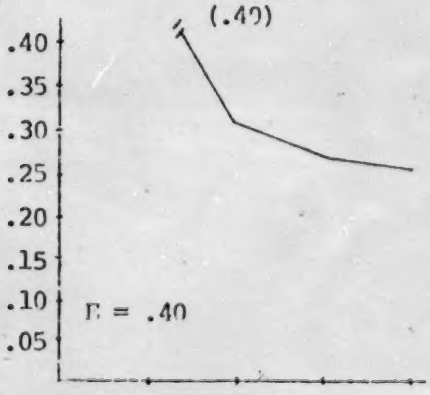
Figure 12

k=20 r=2 n=20









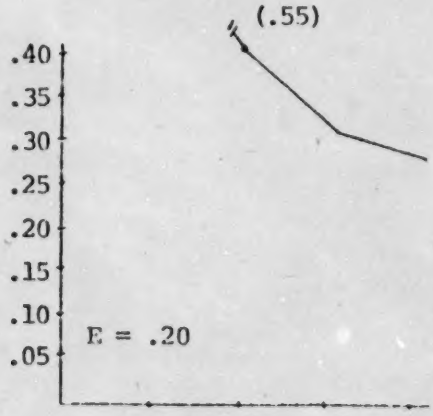
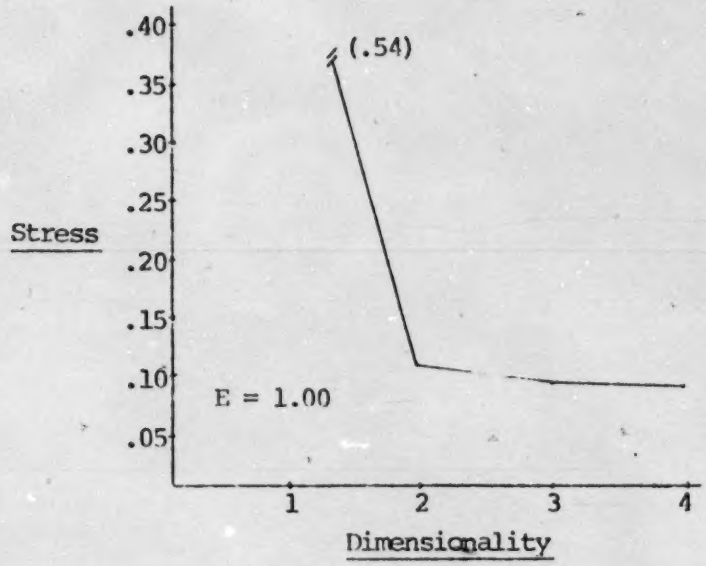
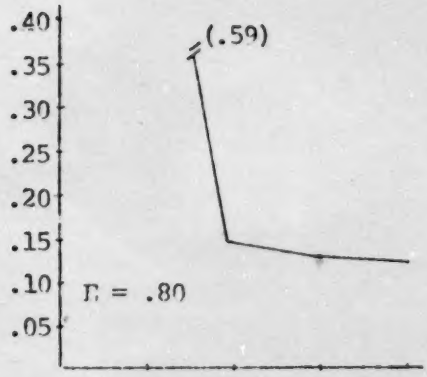
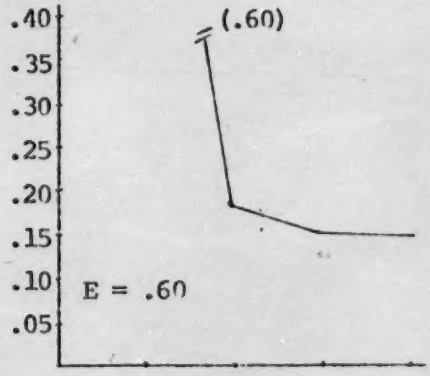


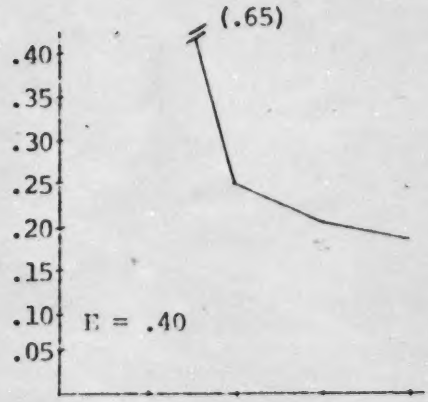
Figure 13

k=20 r=2 n=40









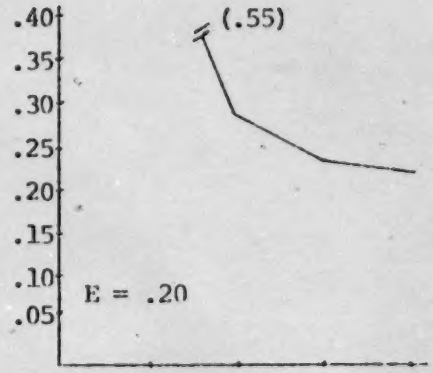
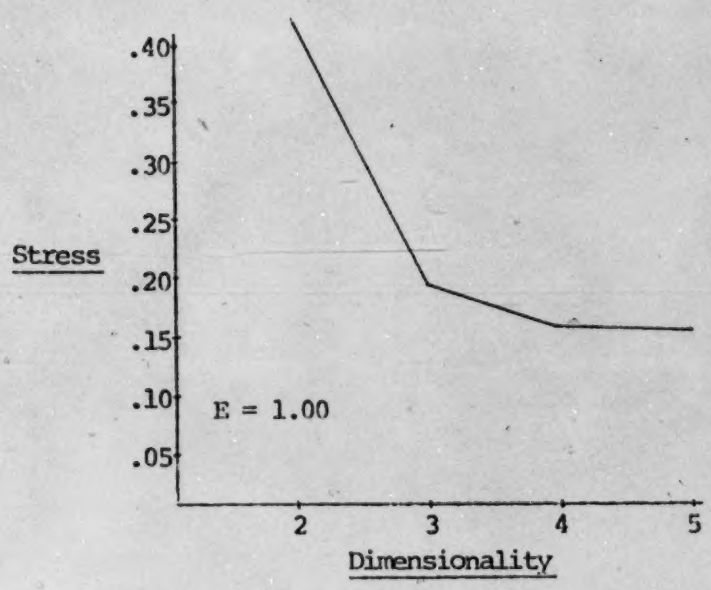
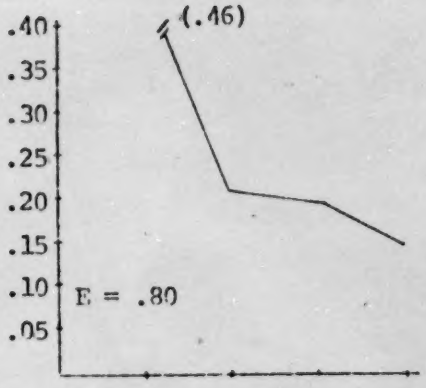
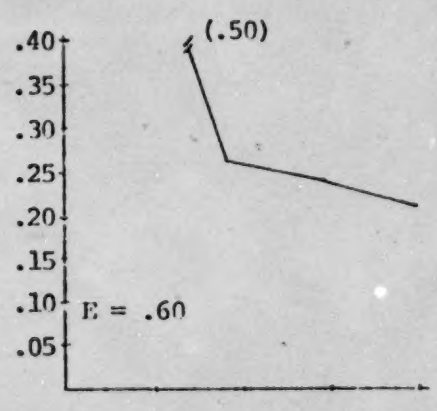


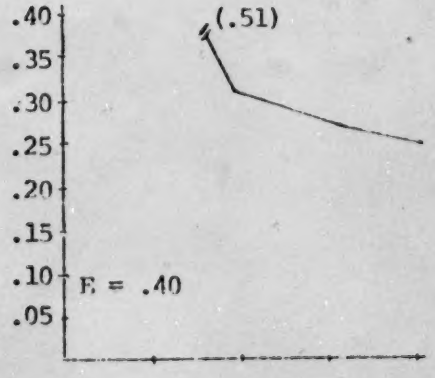
Figure 14

k=20 r=3 n=20









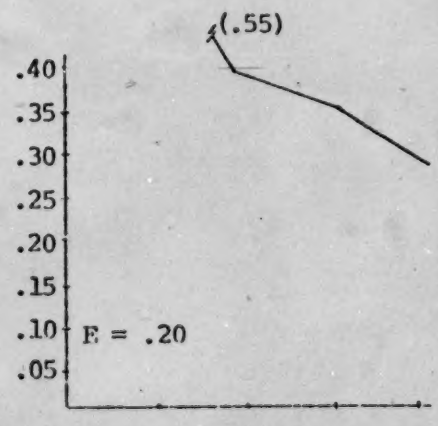
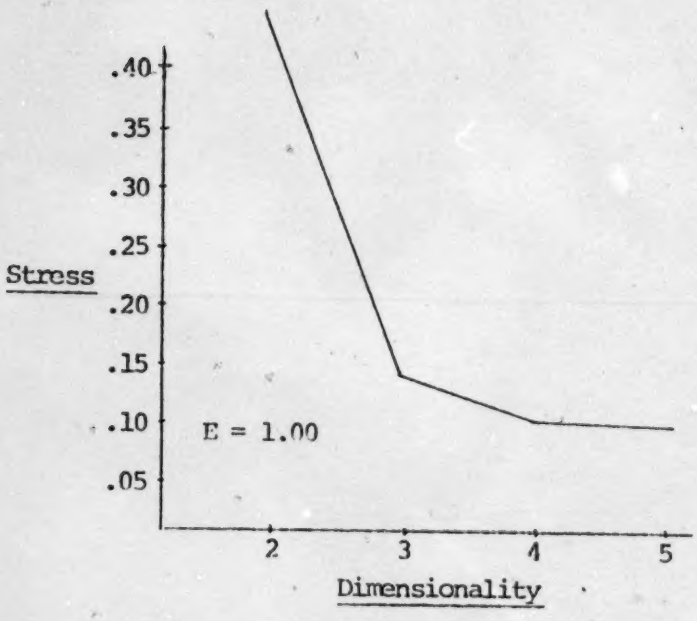
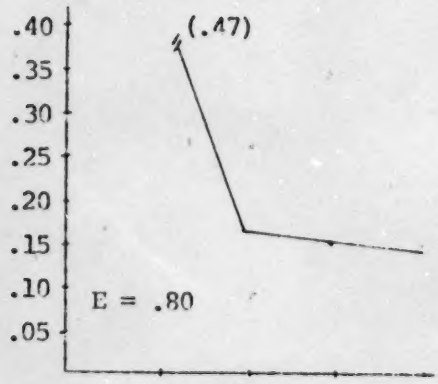
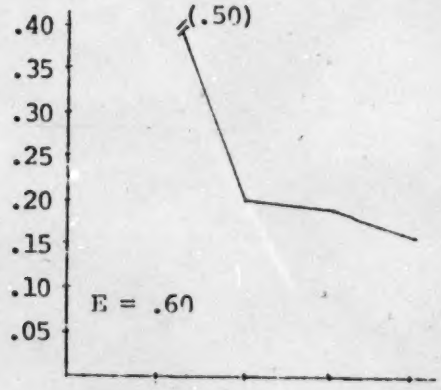


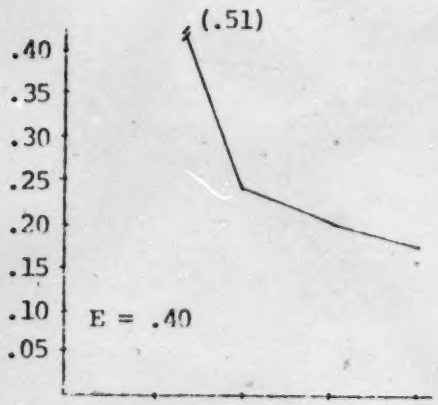
Figure 15

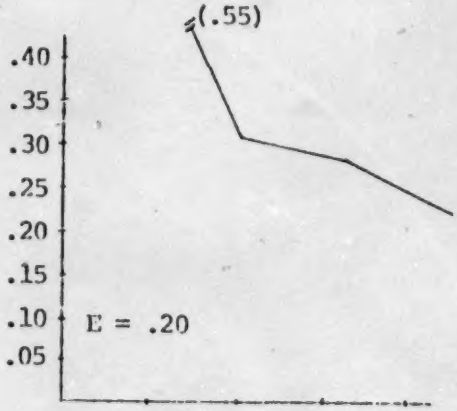
k=20 r=3 n=40











References

Barcikowski, R.S., "A Monte Carlo Study of Item Sampling (Versus Traditional Sampling) for Norm Construction." Journal of Educational Measurement, Vol. IX (1972), pp. 209-215.

Domoney, D.W., I. Spence, and J.C. Ogilvie, "Incomplete Designs for Nonmetric Multidimensional Scaling." The University of Western Ontario Research Bulletin. No. 268 (1973).

Gabriel, R.M., and K.D. Hopkins, "Relative Merits of Univariate ANOVA, Repeated Measures ANOVA, and MANOVA for Research Involving Multiple Criterion Measures." Journal of Special Education, In press (1974).

Glass, G. V, and J.C. Stanley, Statistical Methods in Education and Psychology. Englewood Cliffs, New Jersey, Prentice-Hall, 1970.

Green, P.E., Maheshwari, A., & Rao, V.R., "Self Concept and Brand Preference: An Empirical Application of Multidimensional Scaling." Journal of the Market Research Society, Vol. XI (1969), pp. 343-360.

Kruskal, J.B., "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis." Psychometrika, Vol. XXIX (1964), pp. 1-27. (a)

_____"Nonmetric Multidimensional Scaling: a Numerical Method." Psychometrika, Vol. XXIX (1964), pp. 115-129. (b)

Kruskal, J.B., and F. Carmone, "How to use M-D-SCAL: a Program to do Multidimensional Scaling and Multidimensional Unfolding." Version 5-M. Unpublished manuscript. Murray Hill, New Yersey, Bell Telephone Laboratories, October, 1969.

Lord, F.M., "Estimating Norms by Item Sampling." Educational and Psychological Measurement, Vol. XXII (1962), pp. 259-269.

Millman, J., and G.V Glass, "Rules of Thumb for Writing the ANOVA Table." Journal of Educational Measurement, Vol. IV (1967), pp. 41-51.

Shepard, R.N., A.K. Romney, and S.B. Nerlove, Editors. Multidimensional Scaling Theory and Applications in the Behavioral Sciences. Vol. I/Theory. New York, Seminar Press, 1972.

Sherman, C.R., "Nonmetric Multidimensional Scaling: A Monte Carlo Study of the Basic Parameters." Psychometrika, Vol. XXXVII (1972), pp. 323-355.

Shoemaker, D.M., "An Application of Item-Examinee Sampling to Scaling Attitudes." Journal of Educational Measurement, Vol. VIII (1971), pp. 279-283.

Young, F.W., "Nonmetric Multidimensional Scaling: Recovery of Metric Information." Psychometrika, Vol, XXXV (1970), pp. 455-473.

Young, F.W., and N. Cliff, "Interactive Scaling with Individual Subjects." Psychometrika,Vol. XXXVII (1972), pp. 385-415.